

(12) UK Patent Application (19) GB (11) 2 289 395 (13) A

(43) Date of A Publication 15.11.1995

(21) Application No 9506412.7

(22) Date of Filing 29.03.1995

(30) Priority Data

(31) 08240552

(32) 11.05.1994

(33) US

(71) Applicant(s)

Sony Electronics Inc.

(Incorporated in USA - Delaware)

One Sony Drive, Park Ridge, New Jersey 07656,
United States of America

(72) Inventor(s)

Reuel Andrew Ely

(74) Agent and/or Address for Service

D Young & Co

21 New Fetter Lane, LONDON, EC4A 1DA,
United Kingdom

(51) INT CL⁶

G11B 27/34, H04H 7/00

(52) UK CL (Edition N)

H4R RPX

(56) Documents Cited

GB 2245745 A EP 0245037 A2 WO 86/05025 A1

US 4992892 A

Inspec Abstract No. C9409-5585-007

(58) Field of Search

UK CL (Edition N) H4R RDSD RPVA RPX RSX

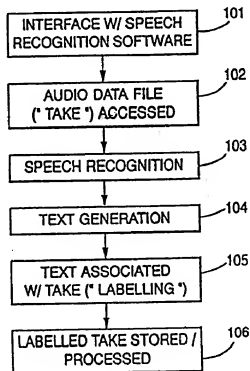
INT CL⁶ G11B 27/34, H04H 7/00

Online: WPI, INSPEC

(54) Method of labelling takes in an audio editing system

(57) In a method of labelling digital audio data corresponding to recorded audio "takes", speech recognition software is employed (102 - 104) to generate text based on a portion of an audio data file corresponding to a dialog take. The generated text is then associated (105) with the audio data file, thereby labelling the file based on its content. The need for monitoring and manual entry of text data in labelling takes is thus eliminated.

FIG. 3



GB 2 289 395 A

FIG. 1

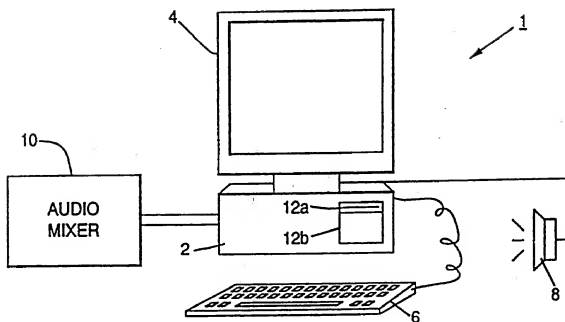


FIG. 2

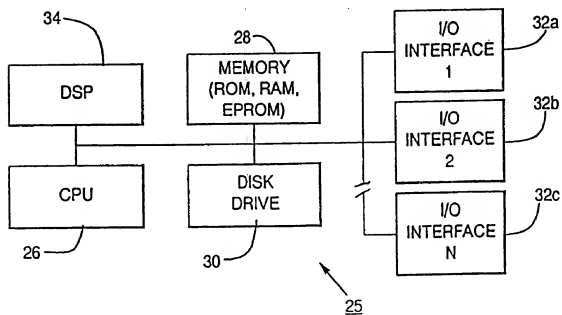
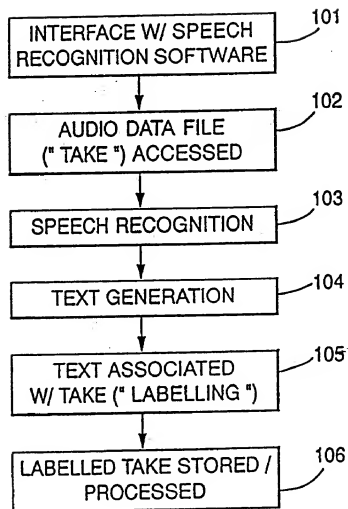


FIG. 3



METHOD OF LABELLING TAKES
IN AN AUDIO EDITING SYSTEM

This invention relates to audio recording. More particularly, it relates to a method of labelling takes in an audio editing system.

5

Audio editing systems are used to perform a variety of functions in conjunction with recorded or live audio. According to one type of system, an editor workstation is employed to process digital data which corresponds to a recorded audio signal. The digital data may be stored in any conventional type of storage medium, such as a disk, or may be stored in memory associated with the workstation. The recorded audio data is commonly referred to as a "take."

Currently, when takes are input to a workstation, a label must be typed in by an operator in order to identify each take. When the take consists of recorded dialog, the label commonly consists of the first few words of recorded dialog. Takes consisting of music might be labelled "MUSIC1," "MUSIC2," and so forth. However, the labelling of takes based on the content of the take requires longer processing than assigning an arbitrary label. This is so because content based labelling consists of inputting the take, listening to the inputted take to formulate a corresponding identifying label, typing the identifying label to be associated with the take, then recording the take with its associated label.

Particularly when this operation is performed repeatedly, it becomes tedious for the operator. This results in fatigue which increases the likelihood of errors in entering appropriate labels, which in turn creates the risk that a stored take will be difficult to

retrieve in the future. More significantly, this operation is costly because a great deal of time must be spent repeatedly typing in labels.

Accordingly, there is a need for an improved method of labelling takes in an audio editing system which is more time efficient and accurate than the conventional manual method.

According to the invention there is provided a method of labelling takes with an audio editing system comprising the following: (a) providing a workstation having memory means and a processor unit associated therewith; (b) programming said workstation with operating system software; (c) accessing with said workstation a digital data file corresponding to a dialog take; (d) interfacing said system software with speech recognition software; (e) subsequent to said interfacing and said accessing, implementing said speech recognition software to translate at least a part of said file into signals representative of text; (f) subsequent to said implementing, associating said signals representative of text with said file; and (g) subsequent to said associating, storing said file and the associated signals representative of text on a storage medium.

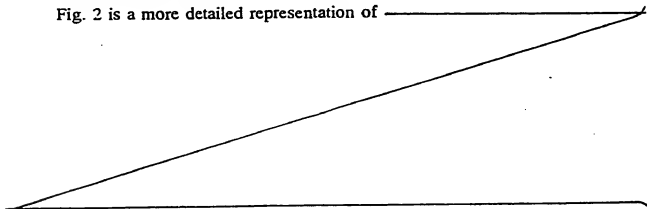
Said workstation may include a central processing unit and a digital signal processor.

Said step of associating may be accomplished by phonetically translating said portion of said file into signals representative of text.

The invention will now be further described, by way of illustrative and non-limiting example, with reference to the accompanying drawings, in which:

Fig. 1 is a representation of an audio editing system for implementing an embodiment of the invention;

Fig. 2 is a more detailed representation of



typical components of a digital audio workstation for implementing an embodiment of the invention; and

Fig. 3 is a flow-chart illustrating an embodiment of the invention.

5

Fig. 1 is a representation of an audio editing system which includes a typical digital audio workstation 1. The digital audio workstation 1 shown comprises, for example, a base unit 2, a monitor 4 and a keyboard 6. The workstation is coupled via an appropriate interface (not shown) to a monitor device, such as a speaker 8. The workstation may additionally be coupled to an audio mixer console according to various techniques known in the art. For example, the workstation may be coupled to the mixer console 10 via a parallel or serial interface through which data may be transferred. Additionally, the workstation typically includes some type of conventional mass storage device, such as a fixed disk drive 12b or floppy disk drive 12a. The storage device is used to store digital data which represents recorded audio signals.

Such a configuration may be used to edit previously recorded audio data stored as digital data. According to the conventional technique, the workstation is appropriately configured with system software to process collections of digital audio data stored, for example, on disk or in memory. In a typical operation, a digital audio data file is accessed and translated into an analog signal which is output to the monitor device. For example, digital audio data might be downloaded from the mixer console to the workstation. The operator then listens to audio signals obtained from the digital audio data using the speaker. The operator then inputs with the keyboard text which identifies the audio data being monitored. Typically this data is stored in ASCII format. The operating software associates the entered text data with the audio data file, and both sets of data

are stored. In this way, each audio data file (each "take") has text data (a "label") stored therewith.

The digital audio workstations now in use possess relatively large processing capabilities. As illustrated in Fig. 2, a digital audio workstation includes a central processing unit 26, various memory areas 28 (for example, ROM, RAM, EPROM), one or more disk drives 30, various input/output interfaces 32a, 32b, and 32c, and one or more digital signal processors (DSPs) 34. The large processing capability offered by such workstations enable convenient editing using conventional graphic display techniques and methods for audio monitoring of recorded takes.

According to the present arrangement, the processing capabilities now available are used to significantly decrease the amount of time necessary to label takes. One embodiment of the invention is described with reference to the flow chart shown as Fig. 3. Of course, variations of this embodiment and other embodiments will be apparent to those skilled in the art. For example, the order of performing the various steps described below may be altered without departing from the scope of the invention.

According to the present embodiment, speech recognition software is made available to the digital audio workstation by any of a number of techniques known in the art. For example, a commercially available program might be installed onto the local storage device, accessed through the system software, and stored in RAM from which it is available to the operator. For example, "Dragon Dictate," commercially available from Dragon Systems of Boston, Massachusetts has the capabilities required for use in conjunction with the system. Of course, the invention is not limited to any specific method of accessing the speech recognition software, nor to any specific speech recognition program.

Once it is made available to the system, the speech recognition software is interfaced through system

software which controls the operation of the digital audio workstation according to conventional techniques, as represented in the first step 101 of Fig. 3. As will be readily apparent to those skilled in the art, the
5 precise steps necessary to achieve this interfacing will vary according to the capabilities of the workstation, and the features of both the speech recognition software and system software employed.

Once the digital audio workstation is properly
10 interfaced with the speech recognition software, stored digital audio data files, that is, takes are then processed using the speech recognition software to obtain text data pursuant to illustrated step 102.

Once stored audio data are accessed in memory
15 or from disk, the speech recognition software is utilized to generate a set of data corresponding to a portion of the take as represented in step 103. Preferably, this operation is accomplished by utilizing any DSP associated with the workstation. For example, the system might be
20 programmed to process the first portion of detected audio monologue and obtaining therefrom text of a predetermined length, such as that corresponding to the first five words of the detected dialog. This step of text generation might be achieved by any conventional
25 technique. For example, according to one common technique, digital audio data is separated into clusters of data which are then converted to text phonetically by use of a stored look-up table.

As shown in step 105, once the text is
30 generated using the speech recognition software, the generated text is associated with the audio data file, that is, the take, from which the text was obtained. The labelled take is then stored, transferred or processed as desired as represented in step 106. For example, the
35 text might be displayed to the operator as a means to verify proper operation of the process.

According to this technique, it is unnecessary to either monitor the stored audio and type in labels as

CLAIMS

1 1. A method of labeling takes in an audio
2 editing system comprising:
3 (a) providing a workstation having memory means
4 and a processor unit associated therewith;
5 (b) programming said workstation with
6 operating system software;
7 (c) accessing with said workstation a digital
8 data file corresponding to a dialog take;
9 (d) interfacing said system software with
10 speech recognition software;
11 (e) subsequent to said interfacing and said
12 accessing, implementing said speech recognition software
13 to translate at least a part of said digital data file
14 into signals representative of text;
15 (f) subsequent to said implementing,
16 associating said signals representative of text with said
17 file; and
18 (g) subsequent to said associating, storing
19 said file and the associated signals representative of
20 text on a storage medium.

1 2. The method of claim 1 wherein said
2 workstation includes a central processing unit and a
3 digital signal processor.

1 3. The method of claim 1 wherein said step of
2 associating is accomplished by phonetically translating
3 said portion of said file into signals representative of
4 text.

1 4. The method of claim 1 wherein said
2 workstation is operatively coupled to an audio mixer
3 console.

1 5. The method of claim 1 wherein said digital
2 data file is downloaded to said workstation from said
3 audio mixer console.

previously required. The method of labelling is thus more time efficient and accurate than the conventional manual method. Moreover, this method may be incorporated readily into a process wherein takes are recorded as digital audio data, downloaded to a video workstation, labelled automatically according to the content of the take and
5 stored.

The foregoing is a detailed description of a preferred embodiment. The scope of the invention, however, is not so limited. Various alternatives will be readily apparent to one of ordinary skill in the art. The invention is only limited by the claims appended hereto.



Application No: GB 9506412.7
Claims searched: 1-6

Examiner: Keith Williams
Date of search: 6 July 1995

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:
UK CI (Ed.N): H4R (RDSD, RPVA,RPX,RSX)
Int CI (Ed.6): G11B 27/34; H04H 7/00
Other: online WPI,INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	GB 2245745 A Sony - see abstract	1
A	EP0245037 A2 Thorn EMI - see abstract	
X	WO86/05025A1 Jostens Learning Systems - see abstract and page 7, lines 31-34; page 28, lines 20-22	
A	US 4992892 Pioneer Electronic - see abstract	1
X	Inspec Abstract No. C9409-5585-007	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

6. A method of labelling takes in an audio editing system, the method being substantially described as herein with reference to the accompanying drawings.